Robust classification of particle tracks for characterization of diffusion and dynamics in fluorescence microscopy

Charles Kervrann, Vincent Briane (PhD student) & Myriam Vimond

Inria Rennes & ENSAI-CREST

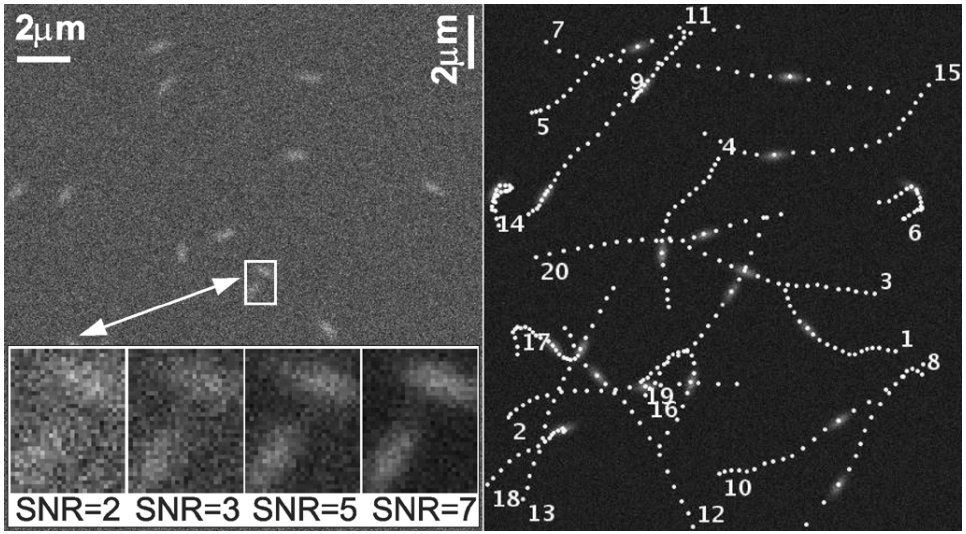
SERPICO Project-Team

http: //www.serpico.rennes.inria.fr

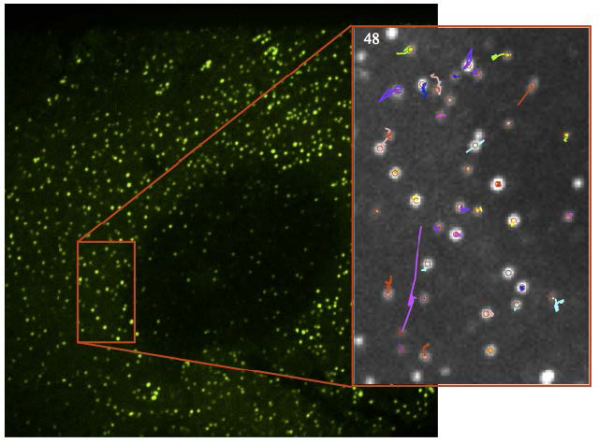
Campus Universitaire de Beaulieu, 35042 Rennes Cedex France

Particle tracking in cell imaging and fluorescence microscopy

Problem: Tracking “key points”, features/descriptors, or random image patches, as long as possible for different signal-to-noise ratios.



Artificial noisy sequences



Real sequences

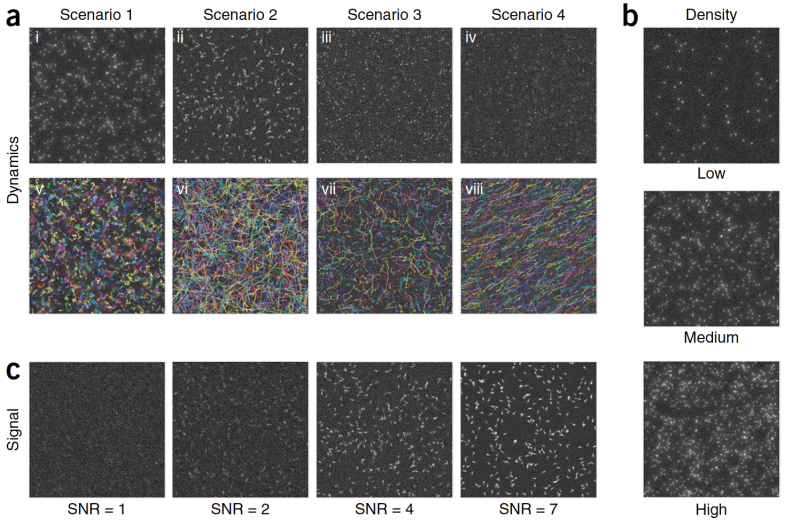
- **Input:** detected/chosen points or patches
- **Matching criterion:** Sum of Squared Differences (SSD), correlation…
- **Output**: tracklets of various objects

Objective comparison of particle tracking methods

Nicolas Chenouard[1–3,25], Ihor Smal[4,5,25], Fabrice de Chaumont[1,25], Martin Maška[6,7,25], Ivo F Sbalzarini[8], Yuanhao Gong[8], Janick Cardinale[8], Craig Carthel[9], Stefano Coraluppi[9], Mark Winter[10], Andrew R Cohen[10], William J Godinez[11,12], Karl Rohr[11,12], Yannis Kalaidzidis[13,14], Liang Liang[15], James Duncan[15], Hongying Shen[16], Yingke Xu[17], Klas E G Magnusson[18], Joakim Jaldén[18], Helen M Blau[19], Perrine Paul-Gilloteaux[20], Philippe Roudot[21], Charles Kervrann[21], François Waharte[20], Jean-Yves Tinevez[22], Spencer L Shorte[22], Joost Willemse[23], Katherine Celler[23], Gilles P van Wezel[23], Han-Wei Dan[24], Yuh-Show Tsai[24], Carlos Ortiz de Solórzano[6], Jean-Christophe Olivo-Marin[1,26] & Erik Meijering[4,5,26]

[1]Institut Pasteur, Unité d'Analyse d'Images Quantitative, Centre National de la Recherche Scientifique Unité de Recherche Associée 2582, Paris, France. [2]Biomedical Imaging Group, École Polytechnique Fédérale de Lausanne, Lausanne, Switzerland. [3]New York University Neuroscience Institute, New York University Medical Center, New York, New York, USA. [4]Department of Medical Informatics, Erasmus University Medical Center, Rotterdam, The Netherlands. [5]Department of Radiology, Erasmus University Medical Center, Rotterdam, The Netherlands. [6]Center for Applied Medical Research, University of Navarra, Pamplona, Spain. [7]Centre for Biomedical Image Analysis, Masaryk University, Brno, Czech Republic. [8]MOSAIC Group, Max Planck Institute of Molecular Cell Biology and Genetics, Dresden, Germany. [9]Compunetix Inc., Monroeville, Pennsylvania, USA. [10]Department of Electrical and Computer Engineering, Drexel University, Philadelphia, Pennsylvania, USA. [11]Department of Bioinformatics and Functional Genomics, Institute of Pharmacy and Molecular Biotechnology, University of Heidelberg, Heidelberg, Germany. [12]Division of Theoretical Bioinformatics, German Cancer Research Center, Heidelberg, Germany. [13]Max Planck Institute of Molecular Cell Biology and Genetics, Dresden, Germany. [14]Belozersky Institute of Physico-Chemical Biology, Moscow State University, Moscow, Russia. [15]Department of Electrical Engineering, Yale University, New Haven, Connecticut, USA. [16]Department of Cell Biology, Yale University, New Haven, Connecticut, USA. [17]Department of Biomedical Engineering, Zhejiang University, Hangzhou, China. [18]Department of Signal Processing, ACCESS Linnaeus Centre, KTH Royal Institute of Technology, Stockholm, Sweden. [19]Baxter Laboratory for Stem Cell Biology, Department of Microbiology and Immunology, Stanford University School of Medicine, Stanford, California, USA. [20]Cell and Tissue Imaging Facility, Institut Curie, Paris, France. [21]Inria Rennes, Bretagne Atlantique, Rennes, France. [22]Plateforme d'Imagerie Dynamique, Imagopole, Institut Pasteur, Paris, France. [23]Molecular Biotechnology Group, Institute of Biology, Leiden University, Leiden, The Netherlands. [24]Department of Biomedical Engineering, Chung Yuan Christian University, Chung Li City, Taiwan, China. [25]These authors contributed equally to this work. [26]These authors jointly directed this work. Correspondence should be addressed to: J.-C.O.-M. (jcolivo@pasteur.fr) or E.M. (meijering@imagescience.org).

RECEIVED 25 MARCH 2013; ACCEPTED 11 DECEMBER 2013; PUBLISHED ONLINE 19 JANUARY 2014; DOI:10.1038/NMETH.2808 NATURE METHODS



Particle tracking is of key importance for quantitative analysis of intracellular dynamic processes from time-lapse microscopy image data. Because manually detecting and following large numbers of individual particles is not feasible, automated computational methods have been developed for these tasks by many groups. Aiming to perform an objective comparison of methods, we gathered the community and organized an open competition in which participating teams applied their own methods independently to a commonly defined data set including diverse scenarios. Performance was assessed using commonly defined measures. Although no single method performed best across all scenarios, the results revealed clear differences between the various approaches, leading to notable practical conclusions for users and developers.

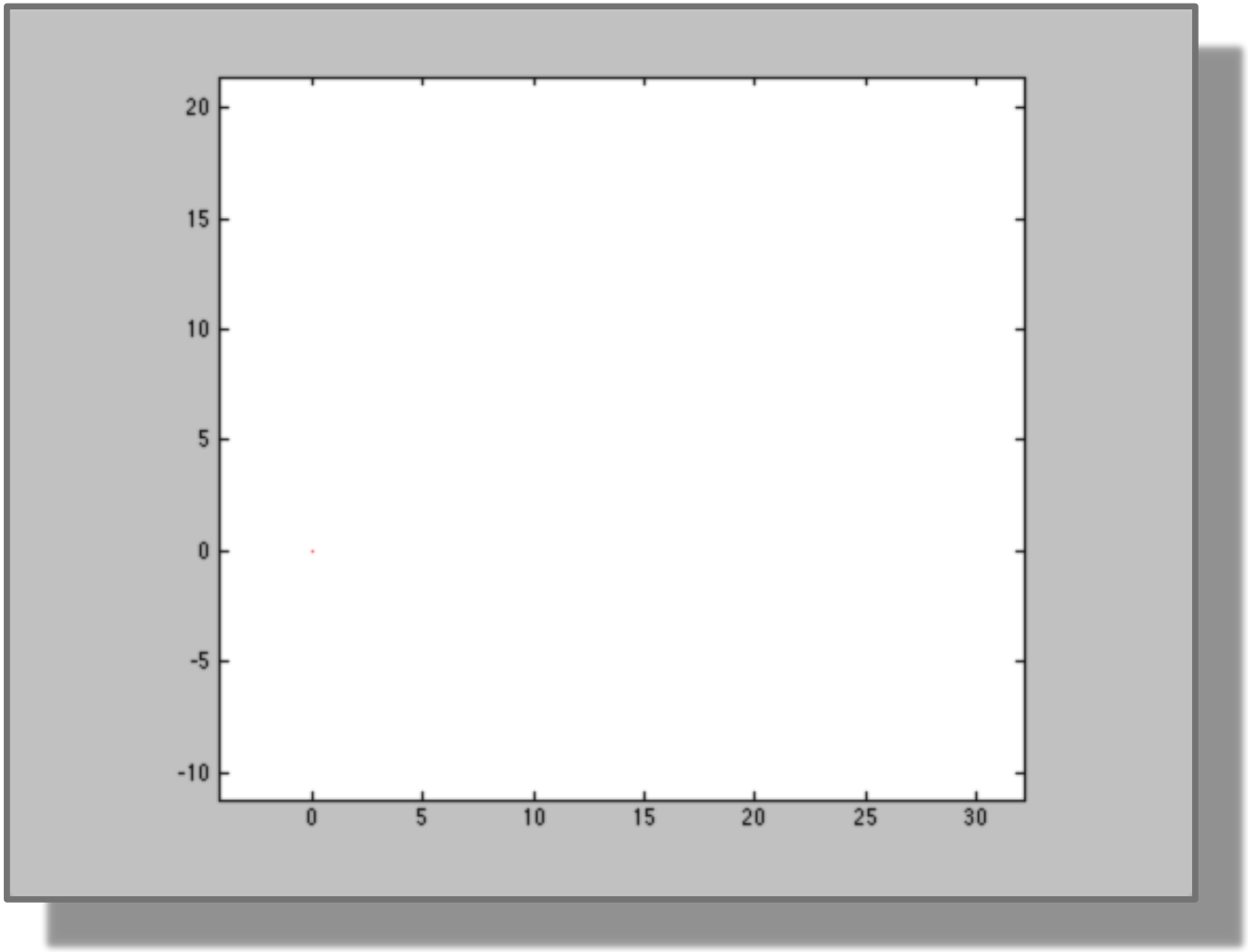
Our objective

Problem statement: Random process and especially **diffusions** can model the trajectories of particles and molecules in live cell imaging.

- We propose a **statistical test** to classify diffusions into 3 groups:
  1. **Free diffusion** (or **Brownian** motion): the particle evolves freely in the domain.
  2. **Superdiffusion**: the particle is transported actively via molecular motors.
  3. **Subdiffusion**: the particle is confined in a domain or evolves in an open but crowded area.

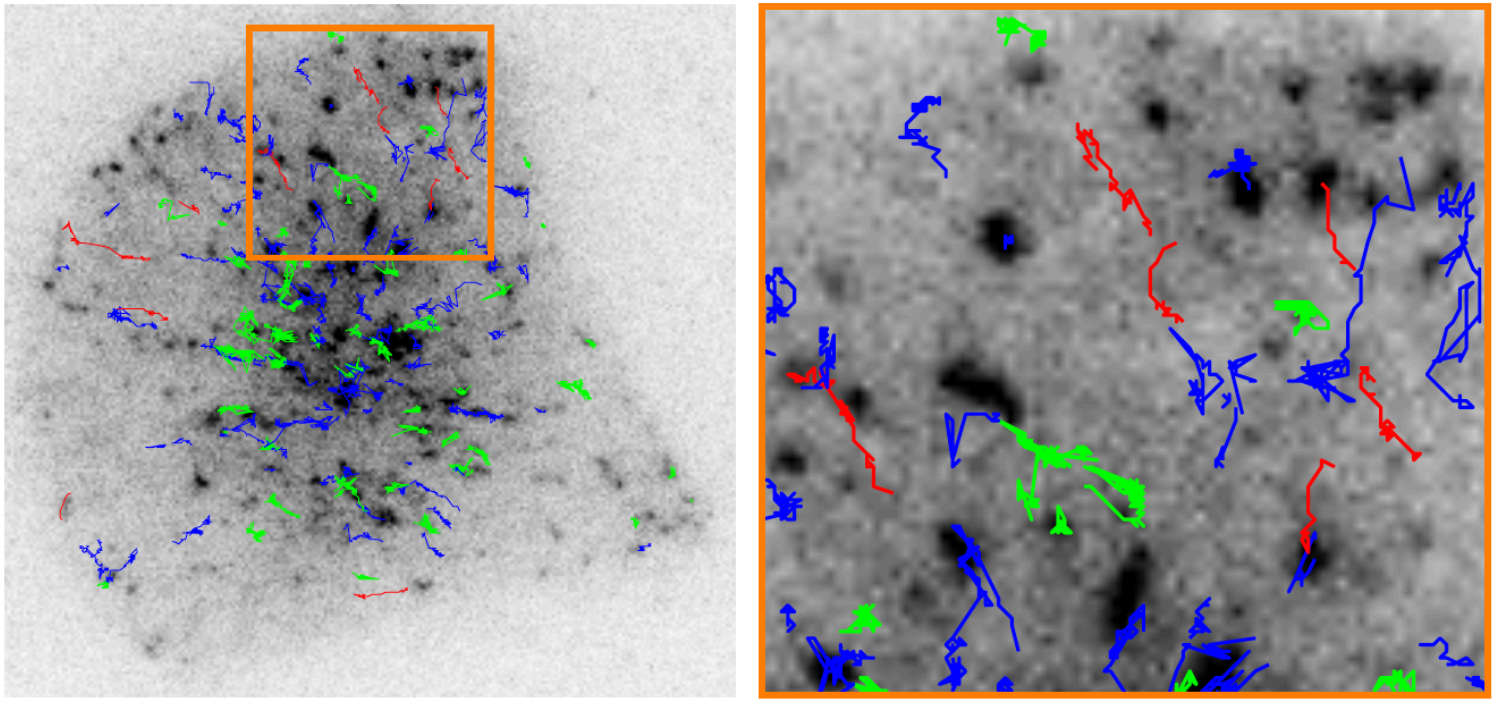
- A commonly used method: **Mean Square Displacement (MSD)**.

A typical simulation



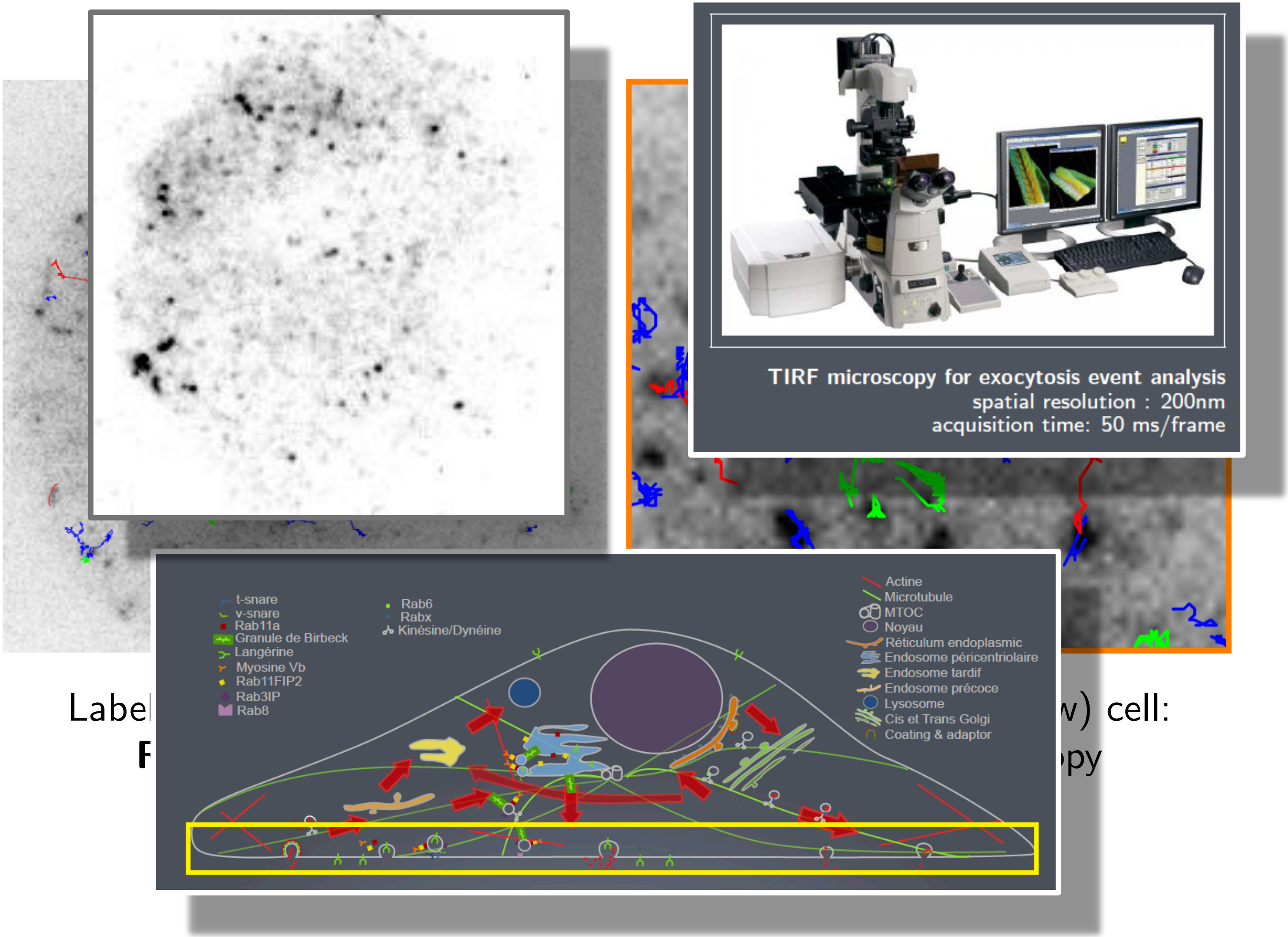
Simulation of **free diffusion**, **superdiffusion** and **subdiffusion**

Data: TIRF microscopy / vesicle trafficking



Labeling of trajectories in a single micro-patterned (crossbow) cell:
Rab11a-GFP protein in TIRF-2D fluorescence microscopy
(courtesy of UMR 144 CNRS-Institut Curie)

Data: TIRF microscopy / vesicle trafficking



Models,
Stochastic Differential Equations
and
Particle Motion in 2D

Definitions of stochastic motions in 2D

- **Brownian motion** $(B_t)_{t>0}$ is a process defined as:

  $$B_0 = 0$$
  $$B_t - B_s \sim \mathcal{N}(0, t-s)$$

  ...continuous and non differentiable paths.

- **Stochastic differential equation (SDE)**:

  $$dX_t = \mu(X_t, t)dt + \sigma(X_t, t)dB_t$$

  where $\mu(X_t, t)$ is the drift (deterministic force) and $\sigma(x, t) = \sigma \mathbf{I}_2$ (random force) is assumed to be isotropic and stationary.

Trajectory model and SDE

- The observed trajectory of a particle is the $(n+1)$-dimensional vector

$$X = (X_{t_0}, X_{t_1}, \ldots, X_{t_n})$$

of successive 2D positions where $X_{t_i} \in \mathbb{R}^2$ and $\Delta t = t_i - t_{i-1}$ is the temporal resolution of the sensor.

- X is generated by the stochastic process $(X_t)_{t_0 \leq t \leq t_n}$ solution of a SDE.

MSD
Mean Square Displacement

Classification of diffusions with MSD

Mean Square Displacement (MSD) to quantify diffusion:

$$MSD(t) = \mathbb{E}(\|X_t - X_0\|_2^2)$$

- **Free diffusion**: $t \mapsto MSD(t)$ is a **linear** function:

$$\mathbb{E}(\|B_t - B_0\|_2^2) = \sigma t.$$

- **Superdiffusion**: $t \mapsto MSD(t)$ **grows faster** than a linear function.
- **Subdiffusion**: $t \mapsto MSD(t)$ **grows slower** than a linear function.

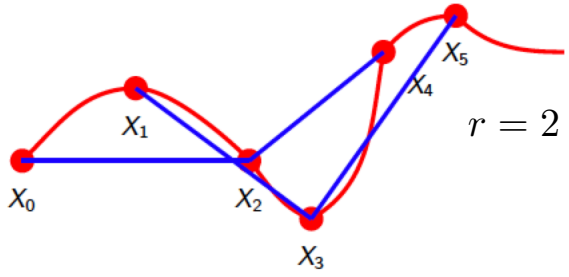
MSD method in practice

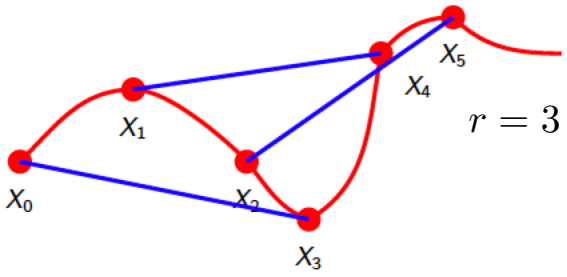
We estimate $t \to \mathrm{MSD}(t)$ from the trajectory X as follows:

$$\widehat{\mathrm{MSD}}(r\Delta t) = \frac{1}{n-r} \sum_{i=1}^{n-r} \|X(t_{i+r}) - X(t_i)\|_2^2$$

When the lag r increases, the performance of the estimator decreases:

- The variance of increments (or distances) increases with r.
- The terms in the sum are correlated (if overlapping).
- The number of terms in the average decreases as r increases.





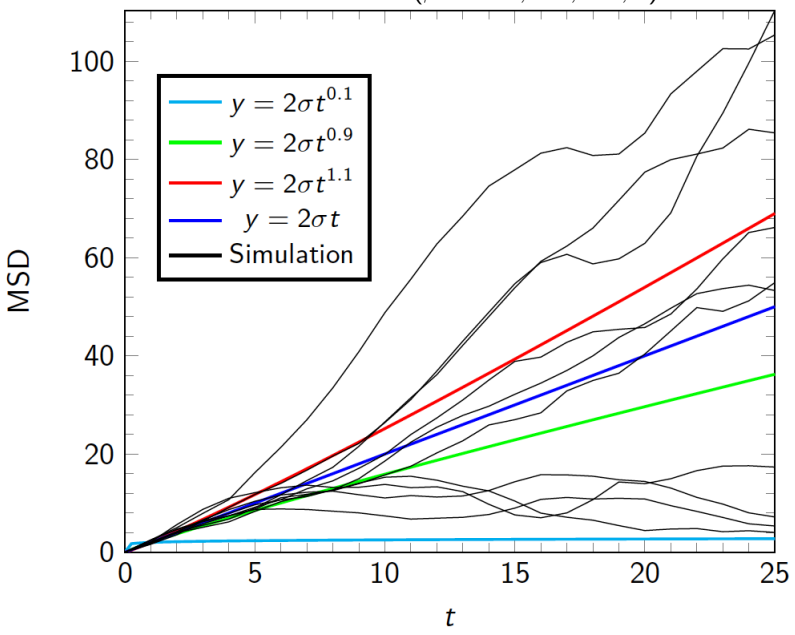
MSD method in practice

Fit $t \mapsto \widehat{\mathrm{MSD}}(t)$ to $t \mapsto Ct^{\beta}$.

Classify according the value of $\beta > 0$ (Saxton 1993, Feder 1996):

- $\beta < 0.1$: **motionless**
- $0.1 < \beta < 0.9$: **subdiffusion**
- $0.9 < \beta < 1.1$: **free diffusion**
- $\beta > 1.1$: **superdiffusion**

Log-log plots of 10 MSD curves of a simulated 2D Brownian motion ($\sigma = 1, n = 30$) and parametric MSD function ($\beta = 0.1, 0.9, 1.1, 9$)



Empirical procedure $\neq$ Statistical procedure

An original
Statistical Test Procedure

A new statistical test for dynamics classification

Our classification procedure is written as a statistical test:

$$H_0 : X_t = \sigma B_t \quad \text{vs} \quad H_1 : (X_t)_{t>0} \text{ is a } \begin{cases} \text{subdiffusion} \\ \text{superdiffusion} \end{cases}$$

- Unlike a conventional binary test, we split H_1 into 2 distinct outcomes.
- A **non parametric** test: under H_1, no parametric assumption on $(X_t)_{t>0}$.

An intuitive test statistic (or measure)

A **measure** to distinguish a **subdiffusion** / **superdiffusion** from **Brownian motion**:

$$S_n = \max_{i=0,\dots,n} \|X(t_i) - X(t_0)\|_2$$

"How far from its initial position did the particle move during the period $[t_n - t_0]$ *?"*

- S_n low : the particle stayed close to X_{t_0} during $[t_0\,,\,t_n]$.
- S_n high : the particle moved far from X_{t_0} during $[t_0\,,\,t_n]$.

Normalized test and motion scaling

Under H_0 the distribution of S_n depends on unknown parameter σ.

- We scale S_n as follows ($\hat{\sigma}$ is a consistent estimator of σ):

$$T_n = \frac{S_n}{\hat{\sigma}\sqrt{t_n - t_0}}$$

Lemma: *Let $\hat{\sigma}$ a consistent estimator of σ such that the distribution of $\hat{\sigma}/\sigma$ does not depend on σ. Then, under H_0, the distribution of T_n does not depend on σ neither on the duration of observation $t_n - t_0$. It depends only on n.*

- The quantile $q_n(\alpha)$ of order α of T_n does not depend on σ.
- T_n has probability $1 - \alpha$ to lie in the region:

$$q_n\left(\frac{\alpha}{2}\right) \leq T_n \leq q_n\left(1 - \frac{\alpha}{2}\right)$$

The test procedure in a few lines

Estimation **off-line** of quantiles $q_n(\alpha)$ (once for all) with **Monte-Carlo simulations** for any trajectory length (n points) and a given α value.

Our test procedure (for an individual trajectory):

1. Estimation of σ: $\widehat{\sigma}^2 = \frac{1}{2n\Delta t} \sum_{i=1}^{n} \|X_{i\Delta t} - X_{(i-1)\Delta t}\|_2^2$
2. Classification of motions according to the **decision rule**:
   - ▷ $T_n \in [q_n(\alpha/2), q_n(1-\alpha/2)]$, then $(X_t)_{t>0}$ is a **Brownian motion**.
   - ▷ $T_n < q_n(\alpha/2)$ then $(X_t)_{t>0}$ is a **subdiffusion**.
   - ▷ $T_n > q_n(1-\alpha/2)$ then $(X_t)_{t>0}$ is a **superdiffusion**.

The test procedure in a few lines

Estimation **off-line** of quantiles $q_n(\alpha)$ (once for all) with **Monte-Carlo simulations** for any trajectory length (n points) and a given α value.

	Trajectory length			
Quantiles ($\alpha = 5\%$)	$n = 10$	$n = 30$	$n = 100$	asymp
$q_n(\alpha/2)$	0.725	0.754	0.785	0.834
$q_n(1-\alpha/2)$	2.626	2.794	2.873	2.940

- $T_n \in [q_n(\alpha/2), q_n(1-\alpha/2)]$, then $(X_t)_{t>0}$ is a **Brownian motion**.
- $T_n < q_n(\alpha/2)$ then $(X_t)_{t>0}$ is a **subdiffusion**.
- $T_n > q_n(1-\alpha/2)$ then $(X_t)_{t>0}$ is a **superdiffusion**.

The test procedure in a few lines

Estimation **off-line** of quantiles $q_n(\alpha)$ (once for all) with **Monte-Carlo simulations** for any trajectory length (n points) and a given α value.

Our test procedure (for an individual trajectory):

1. Estimation of σ: $\widehat{\sigma}^2 = \frac{1}{2n\Delta t} \sum_{i=1}^{n} \|X_{i\Delta t} - X_{(i-1)\Delta t}\|_2^2$
2. Classification of motions according to the **decision rule**:
   - ▷ $T_n \in [q_n(\alpha/2), q_n(1-\alpha/2)]$, then $(X_t)_{t>0}$ is a **Brownian motion**.
   - ▷ $T_n < q_n(\alpha/2)$ then $(X_t)_{t>0}$ is a **subdiffusion**.
   - ▷ $T_n > q_n(1-\alpha/2)$ then $(X_t)_{t>0}$ is a **superdiffusion**.

Experimental Results

Evaluation on synthetic trajectories

Two well-known **parametric superdiffusion and subdiffusion processes**:

- **Superdiffusion**: Brownian motion + constant drift

$$dX_t^i = v_i dt + \sigma dB_t^i \quad i = 1, 2.$$

- **Subdiffusion**: Ornstein-Uhlenbeck process (λ determines the size of the confinement domain)

$$dX_t^i = -\lambda(X_t - \theta_i)dt + \sigma dB_t^i \quad i = 1, 2.$$

A few typical numbers:

- **Length of trajectories**: $n = 30$.
- **Superdiffusion**:

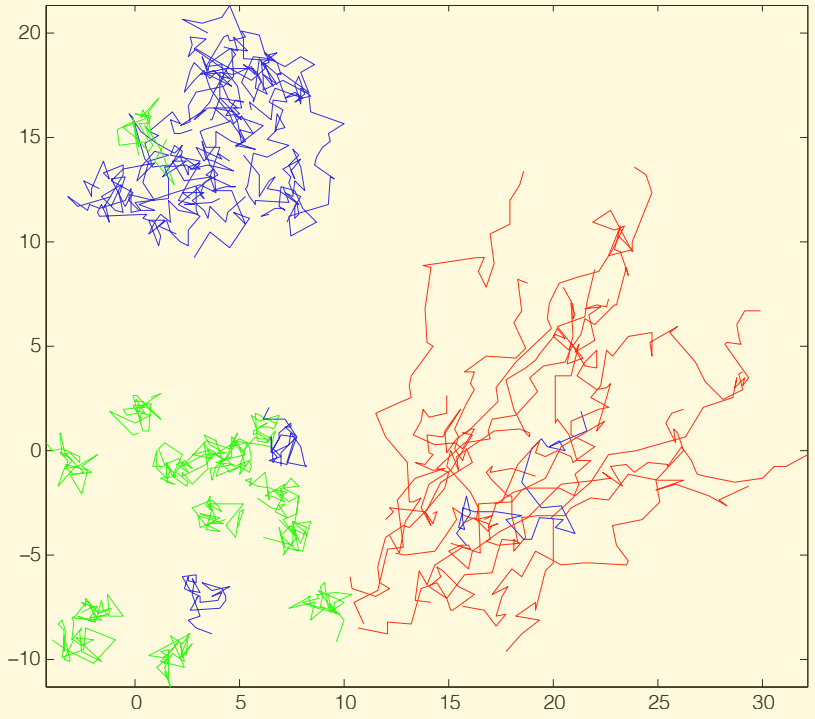
$$\left.\begin{array}{rcl} \Delta t & = & 0.1\,s \\ \sigma^2 & = & 4\,\mu m^2.s^{-1} \\ \|v\| & = & 4.6\,\mu m.s^{-1} \end{array}\right\} \mapsto \frac{\|v\|\sqrt{\Delta t}}{\sigma} = 0.72$$

- **Subdiffusion**:

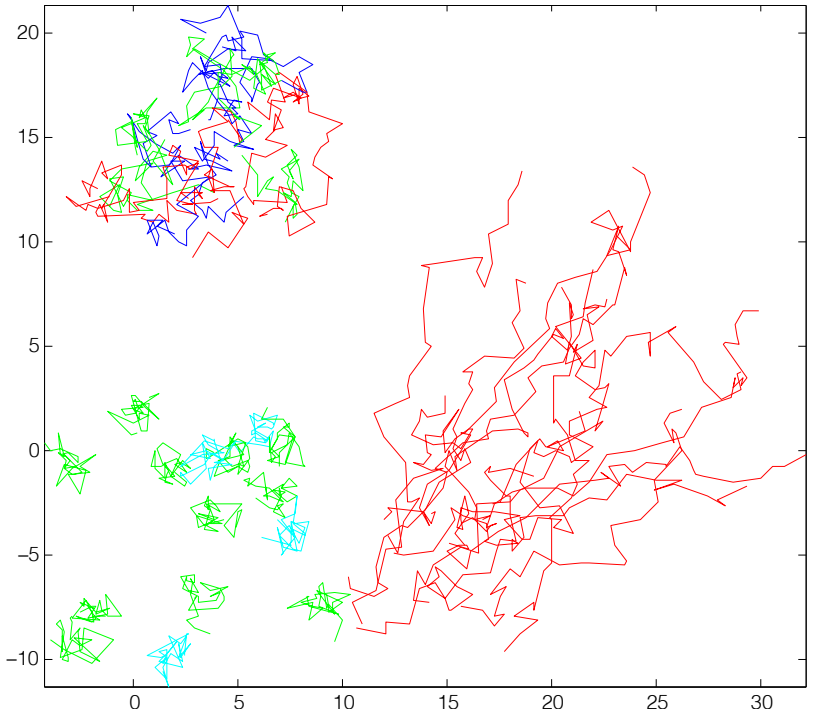
$$\left.\begin{array}{rcl} \Delta t & = & 0.1\,s \\ \lambda & = & 5\,s^{-1} \end{array}\right\} \mapsto \lambda \Delta t = 0.5$$

Evaluation on synthetic trajectories

Our test procedure ($\alpha = 5\%$)



MSD method



Simulated trajectories of **Brownian**, **Brownian with drift** and **Ornstein-Uhlenbeck** (cyan trajectories are labeled as Motionless).

Evaluation on synthetic trajectories

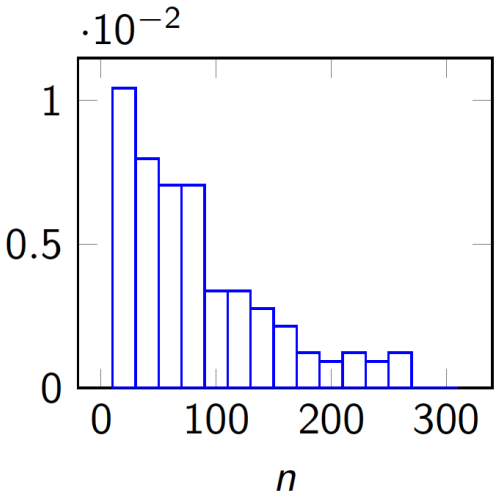
Confusion matrices of our test procedure

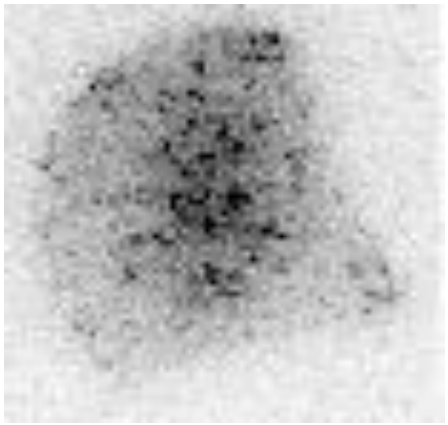
Test label	Brownian	Brownian + drift	Ornstein Uhlenbeck	Brownian	Brownian + drift	Ornstein Uhlenbeck
Ground truth	without noise			with noise		
Brownian	94.6	3	2.7	94.2	1.3	4.5
Brownian + drift	12.7	87.3	0	19.7	80.3	0
Ornstein-Uhlenbeck	26.6	0	73.4	19.8	0	80.2

- Results obtained with $N = 10000$ simulated trajectories of each process with parameters given previously.
- For the noisy case, we set $\sigma_{err} = 0.2$ to get $\sigma\sqrt{\Delta t}/\sigma_{err} = 1$.
- In our results, 1.3% of the simulated **Brownian** trajectories with noise were labeled as **Brownian + drift**.

Evaluation on real 2D-TIRF fluorescence microscopy images (Rab11-GFP protein)

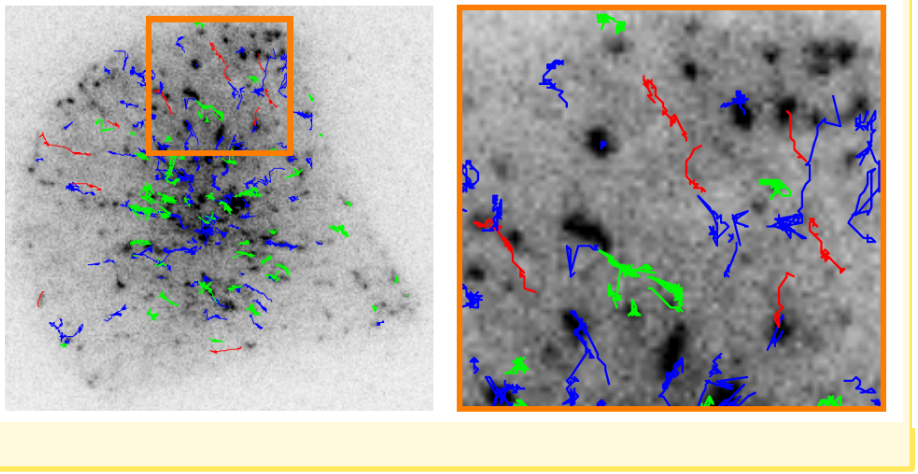
- Sequences of fluorescent images (**TIRF microscopy**) depicting the traffic of **Rab11-GFP protein** in micro-patterned cells (crossbow shape): 600 frames of size 256×240 pixels (1 pixel $=$ 160nm) acquired with $\Delta t = 0.1s$.
- Trajectories computed with the ICY tracker (`icy.bioimageanalysis.org`).
  - N. Chenouard et al., "Multiple Hypothesis Tracking for Cluttered Biological Image Sequences," IEEE Trans. Pattern Anal. Mach. Intell., vol. 35, no. 11, pp. 2736–3750, Nov. 2013
  - F. de Chaumont et al., "Icy: an open bioimage informatics platform for extended reproducible research," Nat. Methods, vol. 9, no. 7, pp. 690–696, Jul. 2012.
- Short trajectories are discarded.



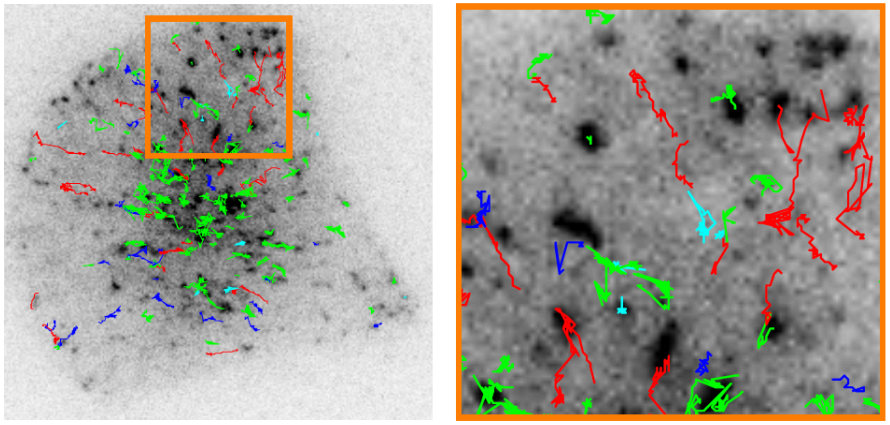


Evaluation on real 2D-TIRF fluorescence microscopy images (Rab11-GFP protein)

Our test procedure



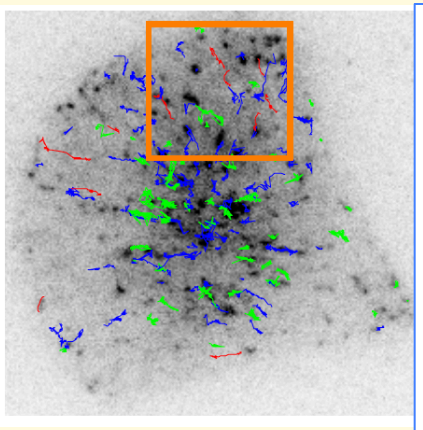
MSD method



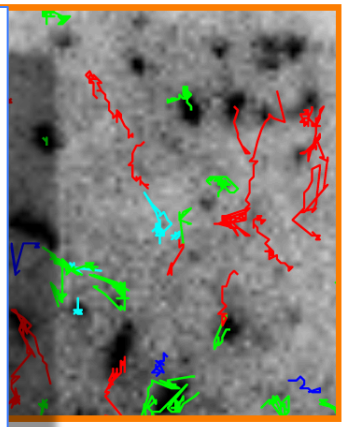
Method Label	Our test procedure	MSD method
Brownian	61%	14%
Subdiffusion	32%	59%
Superdiffusion	7%	20%
Motionless	0	7%

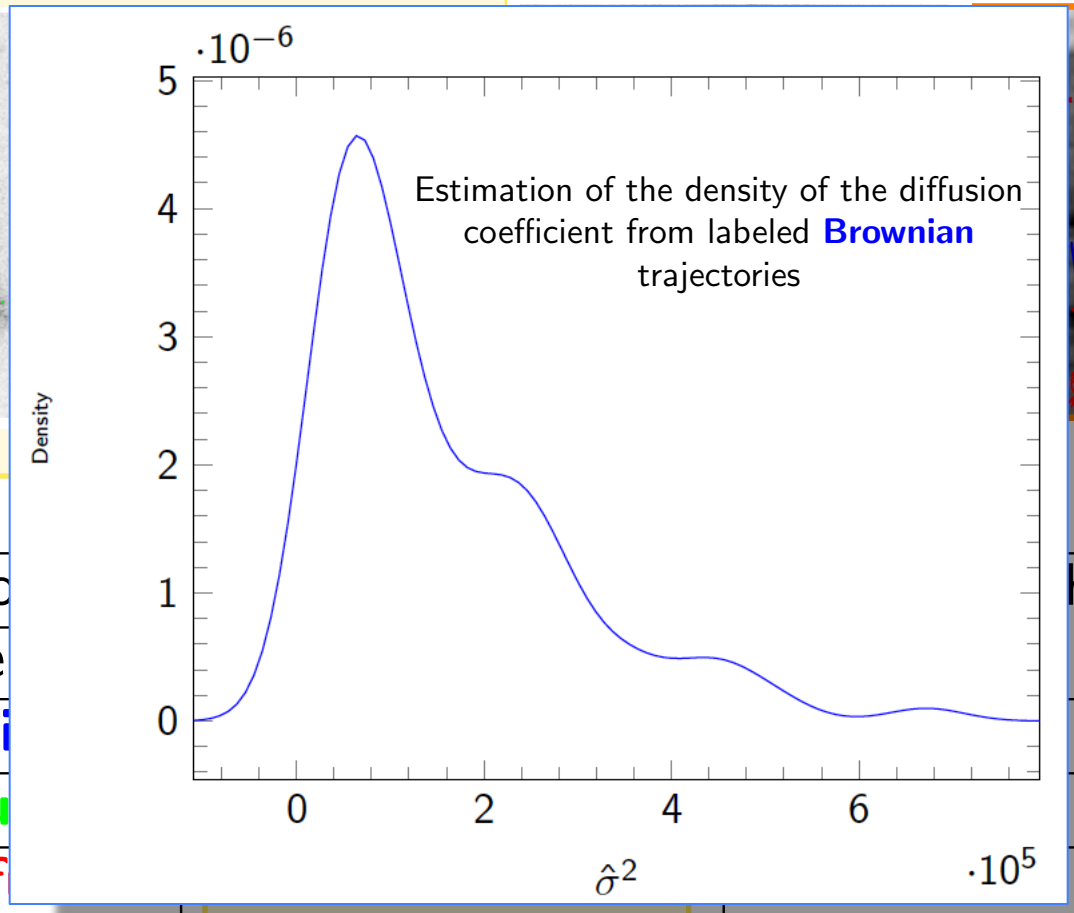
Evaluation on real 2D-TIRF fluorescence microscopy images (Rab11-GFP protein)

Our test procedure



MSD method





Metho		hod
Labe		
Browni		
Subdiffu		
Superdiff		
Motionless	0	7%

Messages to take away

- ▷ Our test procedure is able to reliably classify **subdiffusion** *vs* **Brownian motion** unlike MSD ($n \geq 10$).
- ▷ Our test is **non-parametric** and statistically **consistent**.
- ▷ Our test procedure is **calibrated** to process short and long trajectories: the decision thresholds depend on n.

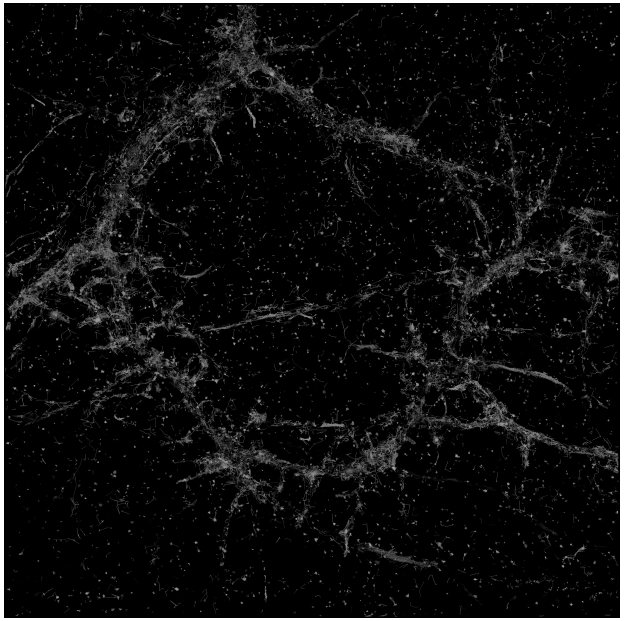
Messages to take away

- ▷ Our test procedure is able to reliably classify **subdiffusion** *vs* **Brownian motion** unlike MSD ($n \geq 10$).
- ▷ Our test is **non-parametric** and statistically **consistent**.
- ▷ Our test procedure is **calibrated** to process short and long trajectories: the decision thresholds depend on n.

Future work

- ▷ Multiple testing for false alarm number reduction: analysis of population of trajectories.
- ▷ Investigation in 3D imaging, super-resolution imaging & **SPT-PALM**.
- ▷ Detection of **motion regime changes** (e.g. exocytosis: transport $\mapsto$ thetering $\mapsto$ docking)

Glutamate receptor subunit 1 of AMPA receptor trajectories (SPT-PALM) moving on the neuronal dendrite surface



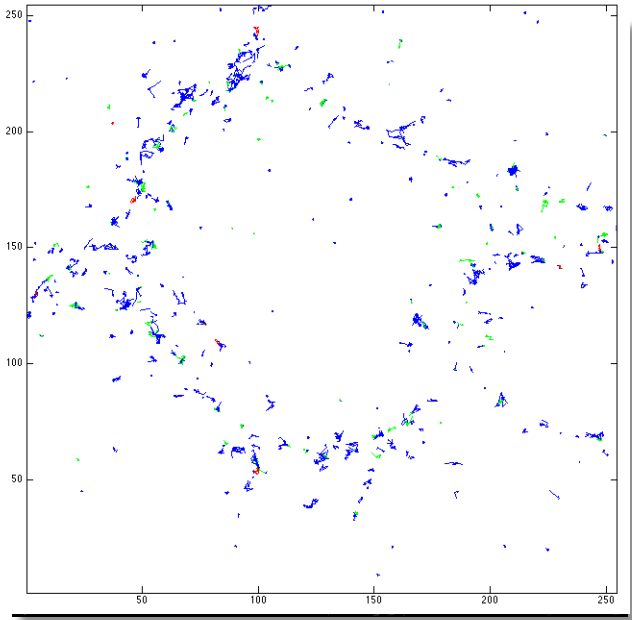
Messages to take away

- ▷ Our test procedure is able to reliably classify **subdiffusion** *vs* **Brownian motion** unlike MSD ($n \geq 10$).
- ▷ Our test is **non-parametric** and statistically **consistent**.
- ▷ Our test procedure is **calibrated** to process short and long trajectories: the decision thresholds depend on n.

Future work

- ▷ Multiple testing for false alarm number reduction: analysis of population of trajectories.
- ▷ Investigation in 3D imaging, super-resolution imaging & **SPT-PALM**.
- ▷ Detection of **motion regime changes** (e.g. exocytosis: transport $\mapsto$ thetering $\mapsto$ docking)

Glutamate receptor subunit 1 of AMPA receptor trajectories (SPT-PALM) moving on the neuronal dendrite surface



Thank you for your attention !

References

Berry, H., & Chaté, H. (2014). *Physical Review E,* **89(2)**, 022708.
Bressloff, P. C., & Newby, J. M. (2013). *Reviews of Modern Physics*, **85(1)**, 135.
Feder, T. J., et al. (1996). *Biophysical J.*, **70(6)**, 2767.
Michalet, X. (2010). *Physical Review E,* **82(4)**, 041914.
Monnier, N., et al. (2012). *Biophysical J.*, **103(3)**, 616-626.
Saxton, M. J., et al. (1997). *Ann. Rev. Biophys. Biomol. Struct.*, **26**, 373-399.
Shaffer, J. P. (1980). *The Annals of Statistics*, 1342-1347.